

The Fragility of Significance in the Hip Arthroscopy Literature

A Systematic Review

Robert L. Parisien, MD, David P. Trofa, MD, Michaela O'Connor, BA, Brock Knapp, BA, Emily J. Curry, BA, Paul Tornetta III, MD, T. Sean Lynch, MD, and Xinning Li, MD

Investigation performed at the Mount Sinai Hospital, New York, Columbia University Irving Medical Center, New York, NY; and Boston University School of Medicine, Boston, Massachusetts

Background: The purpose of the present study was to perform the first examination of the utility of p values and the degree of statistical fragility in the hip arthroscopy literature by applying both the Fragility Index (FI) and the Fragility Quotient (FQ) to dichotomous comparative trials. We hypothesized that dichotomous comparative trials evaluating categorical outcomes in the hip arthroscopy literature are statistically fragile.

Methods: The PubMed and MEDLINE databases were queried from 2008-2018 for comparative studies evaluating dichotomous data in the hip arthroscopy literature. The present analysis included both randomized controlled trials (RCTs) and non-RCTs in which dichotomous data and associated p values were reported. Fragility analysis was performed with use of the Fisher exact test until an alteration of significance was determined.

Results: Of the 5,836 studies screened, 4,156 met the search criteria, with 52 comparative studies included for analysis. One hundred and fifty total outcome events with 33 significant ($p < 0.05$) outcomes and 117 nonsignificant ($p \geq 0.05$) outcomes were identified. The final FI incorporating all 150 outcome events from 52 comparative studies was only 3.5 (interquartile range, 2 to 6), with an associated FQ of 0.032 (interquartile range, 0.017 to 0.063). Twenty-two studies (42.3%) either failed to report loss to follow-up (LTF) data or reported LTF greater than the overall FI of 3.5.

Conclusions: The peer-reviewed hip arthroscopy literature may not be as stable as previously thought, as the sole reliance on a threshold p value has proven misleading. We therefore recommend reporting of the FI and FQ, in conjunction with p values, to aid in the evaluation and interpretation of statistical robustness and quantitative significance in future comparative hip arthroscopy studies.

Hip arthroscopy was developed in 1931 by Burman for joint visualization and began gaining clinical applicability in the 1970s to 1980s as many introduced hip arthroscopy as a technique for diagnosis, management, and surgical intervention for several hip ailments¹⁻³. This procedure has been recently popularized as a joint-preserving procedure for the management of femoroacetabular impingement (FAI), a term that was first coined by Ganz et al. in 2003⁴. Since its introduction, FAI has been rapidly identified as a major cause of hip pain and arthritis⁵. Subsequently, hip arthroscopy has become increasingly utilized, given its favorable outcomes^{6,7}. Evidence-based medicine (EBM) has driven treatment protocols and surgical methods for all fields of medicine, including hip arthroscopy. EBM was first introduced by Cochrane in 1972⁸ and was later expanded by Eddy under the principle of utilizing information on

health and economic outcomes to guide clinical decision-making⁹. The adaptation of EBM has become central to modern medicine but is challenging given the ever-expanding body of published literature. Furthermore, instances of poor data and statistical integrity in orthopaedic research may compromise EBM guidance if they are not properly recognized¹⁰. This is especially true for relatively new techniques such as hip arthroscopy, for which the availability of high-powered studies may be limited.

In the hip arthroscopy literature, dichotomous comparative trials produce the best available evidence to guide clinical decisions, with significance being determined by probability assessment, resulting in either the rejection of, or the failure to reject, the null hypothesis. This method produces an a priori p value threshold set at 0.05, thus representing a 5% likelihood that the difference is due to random chance. Despite its

Disclosure: The **Disclosure of Potential Conflicts of Interest** forms are provided with the online version of the article (<http://links.lww.com/JBJSOA/A332>).

Copyright © 2021 The Authors. Published by The Journal of Bone and Joint Surgery, Incorporated. All rights reserved. This is an open-access article distributed under the terms of the [Creative Commons Attribution-Non Commercial-No Derivatives License 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/) (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

ubiquity, the p value has been met with criticism because of instances in which it may be overvalued without regard for factors such as sample size, loss to follow-up (LTF), or lack of sufficient power¹¹⁻¹³. In these cases, a limited number of event reversals can change study significance. The fragility index (FI), first proposed by Feinstein in 1990 as the “unit fragility,” was developed to address the shortcomings of the p value and is expressed as the number of event reversals required to change study significance¹⁴. A low FI is indicated by only a few event reversals being required to reverse study significance, thus suggesting statistically fragile results. With the FI having been retrospectively applied to the literature, an alarming prevalence of statistical fragility has been identified across several disciplines and subspecialties¹⁵⁻³⁰. Applying the FI in addition to p value analysis provides a much clearer picture of the stability of outcomes. However, the FI is an absolute measure, so it is independent of cohort size. Therefore, Ahmed et al. proposed the Fragility Quotient (FQ) as a means of determining the relative measure of fragility by dividing the FI by the total sample size³¹. Supplementing the p value with the FI and FQ provides a more comprehensive understanding of study stability by accounting for sample size. These stability metrics can aid readers in their critical evaluation of the literature while guiding clinical decision-making through evidence-based principles.

The purpose of the present study was to perform the first examination of the utility of p values and the degree of statistical fragility in the hip arthroscopy literature by applying both the FI and FQ to dichotomous comparative trials. We hypothesized that dichotomous comparative trials evaluating categorical outcomes in the hip arthroscopy literature are statistically fragile.

Materials and Methods

The PubMed and MEDLINE databases were queried from 2008 to 2018 for comparative studies reporting dichotomous data in the hip arthroscopy literature, with utilization of the following search terms: “hip arthroscopy” OR “cam” OR “pincer” OR “labrum” OR “femoroacetabular impingement” OR “FAI” OR (“hip” AND “arthroscopy”) OR (“hip” AND “dysplasia”) OR (“hip” AND “cam”) OR (“hip” AND “pincer”) OR (“hip” AND “labrum”) OR (“hip” AND “femoroacetabular impingement”) OR (“hip” AND “FAI”). Journals included for analysis were *The Journal of Bone and Joint Surgery* (JBJS); *The American Journal of Sports Medicine* (AJSM); *Arthroscopy: The Journal of Arthroscopic and Related Surgery* (Arthroscopy); *Knee Surgery, Sports Traumatology, Arthroscopy* (KSSTA); and the *Journal of Hip Preservation Surgery* (JHPS). These 5 peer-reviewed journals were selected because of their prominence in the published hip arthroscopy literature. Thus, analysis of 11 years of data within these 5 journals provides a representative sample of peer-reviewed research in hip arthroscopy. The analysis included both randomized controlled trials (RCTs) and non-RCTs in which dichotomous data and associated p values were reported. Studies involving cadaveric, animal, in vitro, and non-dichotomous data, along with systematic reviews, were excluded from analysis. Outcome measures were reported as primary,

secondary, or not specified, as specifically stated in each study that was included in the analysis. Outcomes that were reported as significant ($p < 0.05$) or not significant ($p \geq 0.05$) were recorded and analyzed. LTF data were determined and documented. Fragility analysis was performed with use of the Fisher exact test until an alteration of significance was determined (Table I)²⁰. For example, if an outcome was reported as significant, the number of events required to alter the p value to not significant was determined. Similarly, the number of events required to change an outcome from not significant to significant was determined. The resultant numerical value indicates the number required to reverse an outcome event and was recorded as the FI for that event. Additionally, all event reversals were determined and pooled, with the median value representing the FI for the entire study. The FQ was also calculated for each outcome by dividing the FI by the sample size. The total FQ for all outcomes as well as the FQ for RCTs and non-RCTs was determined. Interquartile ranges (IQRs) were determined to aid in the interpretation of the reported variability and dispersion of the data.

Source of Funding

No external funding was acquired in support of this research.

Results

Of the 5,836 studies screened, 4,156 met the search criteria, with 52 comparative studies included in the analysis (Fig. 1). One hundred and fifty total outcome events with 33 significant ($p < 0.05$) outcomes and 117 with nonsignificant ($p \geq 0.05$) outcomes were identified. For the 33 outcomes that were reported as significant, the median number of events required to change significance was only 4 (IQR, 1 to 9) (Table II). The FQ for significant outcomes was 0.025 (IQR, 0.010 to 0.082). For the 117 outcomes that were reported as nonsignificant, the number of events required to change significance was 3 (IQR, 2 to 5). The FQ for nonsignificant outcomes was 0.032 (IQR, 0.017 to 0.060). Therefore, there was no difference in statistical fragility between outcome events reported as significant as compared with those reported as nonsignificant. Of the 150 total outcomes, 34 (22.7%) were primary, 30 (20%) were secondary, and 86 (57.3%) were not specified. No difference was appreciated between primary,

TABLE I Demonstration of Reversal of Significance with a Fragility Index of 1

	Outcome A	Outcome B	P Value
Scenario 1			
Treatment A	1	23	0.04
Treatment B	6	14	
Scenario 2			
Treatment A	2	22	0.11
Treatment B	6	14	

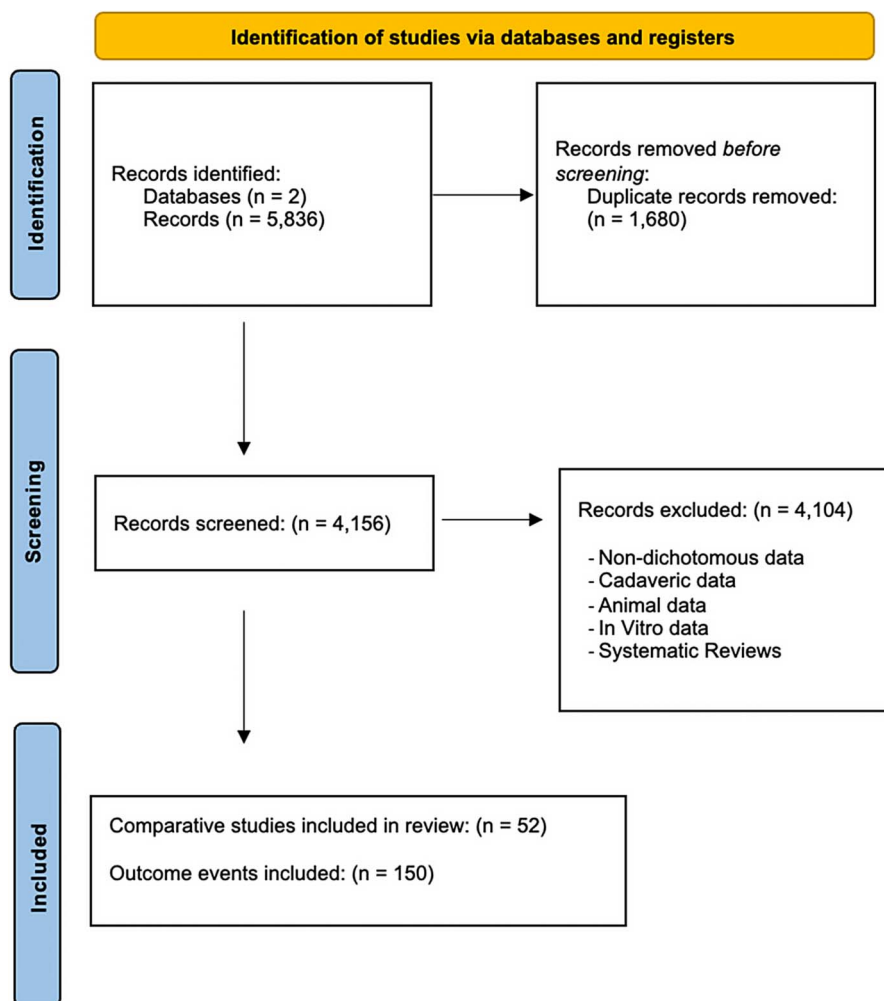


Fig. 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) study identification flowchart.

secondary, and not-specified outcomes, with an FI of 3 (IQR, 2 to 5), 3 (IQR, 2 to 6), and 4 (IQR, 2 to 5), respectively. The associated FQ was nearly identical, with values of 0.028, 0.047, and 0.032, respectively. Further subanalysis by journal did not demonstrate a correlation between study fragility and impact factor (IF), with the least-fragile findings realized in JBJS (IF, 4.578), with an FI of 9 and an associated FQ of 0.071 (Table III). This was followed by Arthroscopy (IF, 4.325), with an FI of 5 and an FQ of 0.068. AJSM (IF, 5.810), JHPS (IF, 1.917), and KSSSTA (IF, 3.210) all demonstrated similar fragility, with an FI of 3 and an FQ of 0.032, 0.028, and 0.004, respectively. A subanalysis of comparative trial types identified a difference between the 141 non-RCT outcomes (FI, 3; IQR, 2 to 5) and the 9 RCT outcomes (FI, 6; IQR, 4.5 to 7). The final FI, incorporating all 150 outcome events from 52 comparative studies, was only 3.5 (IQR, 2 to 6). The final FQ was 0.032 (IQR, 0.017 to 0.063), indicating that the reversal of only 3.2 of 100 outcomes may change the study significance of the included RCTs and non-RCTs. Of the 52 included studies, 19 (36.5%) failed to report LTF data. Three studies (5.8%) reported LTF data greater

than the overall FI of 3.5. Therefore, 42.3% of studies either failed to report LTF data or reported an LTF value that was greater than the overall FI. Subgroup analysis showed that 14.3% of RCTs and 40% of non-RCTs failed to report LTF data.

Discussion

In the present comprehensive evaluation of 52 hip arthroscopy comparative trials and 150 outcome events across 5 leading peer-reviewed orthopaedic journals, we demonstrated substantial fragility, with an overall median FI of only 3.5 and associated median FQ of just 0.032. An FI of 3.5 indicates that reversal of the outcome for just 4 patients would be sufficient to reverse significance. Accounting for sample size, an FQ of just 0.032 indicates a low level of trial stability as only 3.2 of 100 patients is the median number required to reverse significance across all 150 outcome events. Furthermore, 22 (42.3%) of the 52 studies failed to provide LTF data or presented an LTF value that was greater than the overall FI. This suggests that reversal of significance might have been realized by simply maintaining

TABLE II Fragility Data Based on Trial and Outcome Characteristics

Characteristic	Events	Fragility Index (IQR)	Fragility Quotient (IQR)
All trials	150	3.5 (2-6)	0.032 (0.017-0.063)
Outcome			
Primary	34	3 (2-5)	0.028 (0.021-0.055)
Secondary	30	3 (2-6)	0.047 (0.026-0.065)
Not specified	86	4 (2-5)	0.032 (0.010-0.069)
Reported p value			
<0.05	33	4 (1-9)	0.025 (0.010-0.082)
≥0.05	117	3 (2-5)	0.032 (0.017-0.060)
Comparative trial			
RCT	9	6 (4.5-7)	0.098 (0.063-0.127)
Non-RCT	141	3 (2-5)	0.028 (0.016-0.056)

follow-up of all patients in the study. In combination with a low median FI and FQ, these compelling data demonstrate that the hip arthroscopy literature may be more fragile than previously recognized.

A primary purpose of conducting evidence-based research is to improve our collective knowledge base and the quality of clinical care delivery. Information gained with regard to particular treatment strategies and patient outcomes allows for physicians to enter the shared decision-making process armed with objective data. Given that such data are heavily relied on for appropriate clinical management, it is crucial for the strength of significant findings be easily accessible and understood. The current and standard method with which to report significance is p value analysis. A p value of <0.05 confers significance and can be interpreted as a 95% probability that the result was not due to chance alone. In that scenario, one would reject the null hypothesis. However, if the statistical finding is fragile in nature, this may lead to an unintentional type-I (alpha) error. The converse also holds true in that failure to reject the null hypothesis in the setting of a fragile statistical finding may lead to a type-II (beta) error. Thus, the p value should not be utilized as a sole measure of effect. Rather, it should be utilized to aid in the interpretation of evidence, taking into consideration study design and methodological integrity. It is therefore necessary to provide an accurate assessment of a study's statistical fragility in the published literature. As such, inclusion of both the FI and the FQ in the analysis of fragility of comparative trials provides clinicians with a more accurate and comprehensive understanding of trial significance. Although a direct comparison of RCTs and non-RCTs may not be appropriate given the differing integrity of study design, with non-RCTs representing vulnerability to both selection bias and confounding, we identified a difference in the fragility of RCTs as compared with non-RCTs. The FI for RCTs was found to be 6, whereas that for non-RCTs was only 3. In other words, the reversal of only 3 events in non-RCTs is sufficient to provide a reversal of significance, compared with 6

events in RCTs. Additionally, RCTs exhibited an FQ of 0.098, whereas while non-RCTs demonstrated an FQ of 0.028. These findings are consistent with previously published studies in the orthopaedic literature evaluating significance and fragility²²⁻³⁰. Khormae et al. evaluated the fragility of dichotomous outcomes in 17 RCTs in the pediatric orthopaedic literature and identified a median FI of just 3²⁵. Evaniew et al., in an evaluation of 40 RCTs in the spine surgery literature, reported a median FI of only 2, with 75% of the trials demonstrating an FI of ≤3²². Parisien et al. further investigated the statistical stability of 102 comparative RCTs and non-RCTs in the sports medicine literature and identified an FI of only 5 across 339 outcome events²⁶. Khan et al., in a study of 48 primary outcomes in RCTs in the sports medicine literature, reported an even more fragile FI of only 2²⁴. Furthermore, Parisien et al., in an evaluation of 775 outcome events across 80 RCTs and 118 non-RCTs in the orthopaedic trauma literature, identified an FI and FQ of just 5 and 0.046, respectively²⁷. Forrester et al., in an examination of 23 studies with 48 outcome events in the orthopaedic oncology literature, identified an overall median FI of 4, with a median FI for significant outcomes of only 2²³. Parisien et al., in 2 recent fragility analyses of RCTs in the cartilage restoration and rotator cuff literature, identified an overall FI of 4 and 4, respectively, as well as an FQ of 0.067 and 0.092, respectively^{29,30}. The FI values for those orthopaedic studies align closely with that of our current evaluation of the hip arthroscopy literature. A number of studies evaluated additional statistical fragility correlates. Three fragility studies^{22,24,25} found that an increasing FI correlated significantly with smaller (more significant) p values, and 3 studies^{22,23,25} reported a positive correlation between FI and sample size. Interestingly, several fragility studies reported on outcomes resulting in an FI of 0, meaning the reversal of significance was determined by simply re-calculating the p value with an alternative statistical test. Khormae et al. identified 3 articles (17.6%) with an FI of 0²⁵. Similarly, Evaniew et al. reported that 8 (20%) of the 40 outcomes that they assessed resulted in an FI of 0 following their own p value analysis²². Khan et al., in a study of the sports medicine literature, reported that an FI of 0 was identified for 8 outcomes (16.6%), leading the authors to report that "outcomes became nonsignificant when we recalculated the p value using the 2-sided Fisher exact test."²⁴ Several studies further evaluated the effect of the number of patients with LTF on the resulting significance. Khormae et al.²⁵, in an evaluation of 17

TABLE III Fragility Data Based on Journal Impact Factor

Journal	Impact Factor	FI	FQ
JBJS	4.578	9	0.071
Arthroscopy	4.325	5	0.068
AJSM	5.810	3	0.032
JHPS	1.917	3	0.028
KSSTA	3.210	3	0.004

pediatric RCTs, reported that only 2 studies actually included LTF data, with 1 study revealing that the number of patients LTF was greater than the resultant FI²⁵. This finding would suggest the potential reversal of study significance by simply maintaining follow-up. Similarly, Evaniew et al., in a comprehensive evaluation of the spine literature, found that the FI was less than or equal to the LTF value for 26 outcomes (65%)²². This pattern persisted in the sports medicine literature, with Khan et al.²⁴ identifying 23 outcomes (48%) with an LTF value that was greater than or equal to the FI. Additionally, in an evaluation of the sports medicine literature, Parisien et al.²⁶ reported that the average LTF value (7.9) was greater than the overall FI of 5. Furthermore, in an evaluation of the orthopaedic oncology literature, Forrester et al. found that 60% of the outcomes had an FI value that was less than or equal to the LTF value²³. Parisien et al., in a recent fragility analysis evaluating the cartilage restoration literature, found that 15.8% of studies either did not report LTF data or reported an LTF value that was greater than the FI²⁹. Additionally, in a systematic review and meta-analysis of RCTs evaluating the use of platelet-rich plasma in rotator cuff surgery, Parisien et al. revealed that, of the studies reporting LTF data, 30.2% reported an LTF value that was greater than the FI³⁰.

The present study is the first to provide a detailed analysis of significance in the hip arthroscopy literature. Our findings further demonstrate the lack of correlation between journal impact factor and degree of fragility, thus emphasizing the importance of including measures such as the FI and FQ to provide additional context to reported p values. Additionally,

the present study includes an analysis of both primary and secondary outcomes for a more comprehensive and accurate FI and FQ analysis.

Conclusions

The peer-reviewed hip arthroscopy literature may not be as stable as previously thought, as the utilization of a threshold p value has proven misleading. We therefore recommend reporting of the FI, FQ, and p value to aid in the evaluation and interpretation of statistical robustness and quantitative significance in future comparative hip arthroscopy studies. ■

Robert L. Parisien, MD¹
David P. Trofa, MD²
Michaela O'Connor, BA²
Brock Knapp, BA³
Emily J. Curry, BA³
Paul Tornetta III, MD³
T. Sean Lynch, MD²
Xinning Li, MD³

¹Mount Sinai Hospital, New York, NY

²Columbia University Irving Medical Center, New York, NY

³Boston University School of Medicine, Boston, Massachusetts

Email for corresponding author: Robert.Parisien@MountSinai.org

References

- Burman MS. Arthroscopy or the direct visualization of joints: an experimental cadaver study. 1931. *Clin Orthop Relat Res.* 2001 Sep;(390):5-9.
- Eriksson E, Arvidsson I, Arvidsson H. Diagnostic and operative arthroscopy of the hip. *Orthopedics.* 1986 Feb;9(2):169-76.
- Johnson LL. Diagnostic and surgical arthroscopy. *Clin Symp.* 1982;34(3):2-32.
- Ganz R, Parvizi J, Beck M, Leunig M, Nötzli H, Siebenrock KA. Femoroacetabular impingement: a cause for osteoarthritis of the hip. *Clin Orthop Relat Res.* 2003 Dec;(417):112-20.
- Byrd JW. Hip arthroscopy: patient assessment and indications. *Instr Course Lect.* 2003;52:711-9.
- Philippon MJ, Schenker ML. Arthroscopy for the treatment of femoroacetabular impingement in the athlete. *Clin Sports Med.* 2006 Apr;25(2):299-308, ix.
- Philippon MJ, Stubbs AJ, Schenker ML, Maxwell RB, Ganz R, Leunig M. Arthroscopic management of femoroacetabular impingement: osteoplasty technique and literature review. *Am J Sports Med.* 2007 Sep;35(9):1571-80.
- Cochrane AL. Effectiveness and efficiency: Random reflections on health services. London: Nuffield Provincial Hospitals Trust, 1972.
- Eddy DM. Clinical decision making: from theory to practice. Connecting value and costs. Whom do we ask, and what do we ask them? *JAMA.* 1990 Oct 3;264(13):1737-9.
- Parsons NR, Hiskens R, Price CL, Achten J, Costa ML. A systematic survey of the quality of research reporting in general orthopaedic journals. *J Bone Joint Surg Br.* 2011 Sep;93(9):1154-9.
- Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. *JAMA.* 2005 Jul 13;294(2):218-28.
- Pocock SJ. Current issues in the design and interpretation of clinical trials. *Br Med J (Clin Res Ed).* 1985 Jan 5;290(6461):39-42.
- Sterne JA, Davey Smith G. Sifting the evidence-what's wrong with significance tests? *BMJ.* 2001 Jan 27;322(7280):226-31.
- Feinstein AR. The unit fragility index: an additional appraisal of "statistical significance" for a contrast of two proportions. *J Clin Epidemiol.* 1990;43(2):201-9.
- Docherty KF, Campbell RT, Jhund PS, Petrie MC, McMurray JJV. How robust are clinical trials in heart failure? *Eur Heart J.* 2017 Feb 1;38(5):338-45.
- Matics TJ, Khan N, Jani P, Kane JM. The Fragility Index in a Cohort of Pediatric Randomized Controlled Trials. *J Clin Med.* 2017 Aug 14;6(8):6.
- Chavalaras D, Wallach JD, Li AH, Ioannidis JP. Evolution of Reporting P Values in the Biomedical Literature, 1990-2015. *JAMA.* 2016 Mar 15;315(11):1141-8.
- Shochet LR, Kerr PG, Polkinghorne KR. The fragility of significant results underscores the need of larger randomized controlled trials in nephrology. *Kidney Int.* 2017 Dec;92(6):1469-75.
- Shen C, Shamsudeen I, Farrokhvar F, Sabri K. Fragility of Results in Ophthalmology Randomized Controlled Trials: A Systematic Review. *Ophthalmology.* 2018 May;125(5):642-8.
- Walsh M, Srinathan SK, McAuley DF, Mrkobrada M, Levine O, Ribic C, Molnar AO, Dattani ND, Burke A, Guyatt G, Thabane L, Walter SD, Pogue J, Devereaux PJ. The statistical significance of randomized controlled trial results is frequently fragile: a case for a Fragility Index. *J Clin Epidemiol.* 2014 Jun;67(6):622-8.
- Ridgeon EE, Young PJ, Bellomo R, Mucchetti M, Lembo R, Landoni G. The Fragility Index in Multicenter Randomized Controlled Critical Care Trials. *Crit Care Med.* 2016 Jul;44(7):1278-84.
- Evaniew N, Files C, Smith C, Bhandari M, Ghert M, Walsh M, Devereaux PJ, Guyatt G. The fragility of statistically significant findings from randomized trials in spine surgery: a systematic survey. *Spine J.* 2015 Oct 1;15(10):2188-97.
- Forrester LA, Jang E, Lawson MM, Capi A, Tyler WK. Statistical Fragility of Surgical and Procedural Clinical Trials in Orthopaedic Oncology. *J Am Acad Orthop Surg Glob Res Rev.* 2020 Jun 1;4(6):e19.00152.
- Khan M, Evaniew N, Gichuru M, Habib A, Ayeni OR, Bedi A, Walsh M, Devereaux PJ, Bhandari M. The Fragility of Statistically Significant Findings From Randomized Trials in Sports Surgery: A Systematic Survey. *Am J Sports Med.* 2017 Jul;45(9):2164-70.
- Khormaee S, Choe J, Ruzbarsky JJ, Agarwal KN, Blanco JS, Doyle SM, Dodwell ER. The Fragility of Statistically Significant Results in Pediatric Orthopaedic Randomized Controlled Trials as Quantified by the Fragility Index: A Systematic Review. *J Pediatr Orthop.* 2018 Sep;38(8):e418-23.

- 26.** Parisien RL, Trofa DP, Dashe J, Cronin PK, Curry EJ, Fu FH, Li X. Statistical Fragility and the Role of P Values in the Sports Medicine Literature. *J Am Acad Orthop Surg.* 2019 Apr 1;27(7):e324-9.
- 27.** Parisien RL, Dashe J, Cronin PK, Bhandari M, Torretta P 3rd. Statistical Significance in Trauma Research: Too Unstable to Trust? *J Orthop Trauma.* 2019 Dec; 33(12):e466-70.
- 28.** Svantesson E, Hamrin Senorski E, Danielsson A, Sundemo D, Westin O, Ayeni OR, Samuelsson K. Strength in numbers? The fragility index of studies from the Scandinavian knee ligament registries. *Knee Surg Sports Traumatol Arthrosc.* 2020 Feb;28(2):339-52.
- 29.** Parisien RL, Constant M, Saltzman BM, Popkin CA, Ahmad CS, Li X, Trofa DP. The Fragility of Statistical Significance in Cartilage Restoration of the Knee: A Systematic Review of Randomized Controlled Trials. *Cartilage.* 2021 May 10: 19476035211012458.
- 30.** Parisien RL, Ehlers C, Cusano A, Torretta P 3rd, Li X, Wang D. The Statistical Fragility of Platelet-Rich Plasma in Rotator Cuff Surgery: A Systematic Review and Meta-analysis. *Am J Sports Med.* 2021 Mar 1:363546521989976.
- 31.** Ahmed W, Fowler RA, McCredie VA. Does Sample Size Matter When Interpreting the Fragility Index? *Crit Care Med.* 2016 Nov;44(11):e1142-3.